

# An Efficient Pest Classification In Smart Agriculture Using Transfer Learning

Tuan T. Nguyen<sup>1</sup>, Quoc-Tuan Vien<sup>2</sup>, Harin Sellahewa<sup>1</sup>

<sup>1</sup>School of Computing, University of Buckingham, United Kingdom. Email: {Tuan.Nguyen, Harin.Sellahewa}@buckingham.ac.uk.

<sup>2</sup>Faculty of Science and Technology, Middlesex University, United Kingdom. Email: Q.Vien@mdx.ac.uk.

## Abstract

To this day, agriculture still remains very important and plays considerable role to support our daily life and economy in most countries. It is the source of not only food supply, but also providing raw materials for other industries, e.g. plastic, fuel. Currently, farmers are facing the challenge to produce sufficient crops for expanding human population and growing in economy, while maintaining the quality of agriculture products. Pest invasions, however, are a big threat to the growth crops which cause the crop loss and economic consequences. If they are left untreated even in a small area, they can quickly spread out other healthy area or nearby countries. A pest control is therefore crucial to reduce the crop loss. In this paper, we introduce an efficient method basing on deep learning approach to classify pests from images captured from the crops. The proposed method is implemented on various EfficientNet and shown to achieve a considerably high accuracy in a complex dataset, but only a few iterations are required in the training process.

Received on 04 December 2020; accepted on 19 January 2021; published on 26 January 2021

**Keywords:** Deep learning, Transfer Learning, Object Classification, Smart Agriculture

Copyright © 2021 Tuan T. Nguyen *et al.*, licensed to EAI. This is an open access article distributed under the terms of the [Creative Commons Attribution license](#), which permits unlimited use, distribution and reproduction in any medium so long as the original work is properly cited.

doi:10.4108/eai.26-1-2021.168227

## 1. Introduction

Human civilisation has changed dramatically. Agriculture, however, still remains very important for every country. Indeed, many raw materials like cotton, sugar, wood, or palm oil, come from agriculture. These materials are essential to major industries such as the manufacturing of pharmaceuticals, diesel fuel, plastic, and more. The most important aspect of agriculture is that it is the source of the world's food supply.

The United Nation predicted the human population will reach 10 billion by the year 2050. Therefore, the amount of food, we produce now, need to be doubled to accommodate this change. Not only increasing the quantity, but also maintaining the quality is a big challenge for farmers in the next generations.

One of the main sources causing damage to growing crops is pest invasions. They can eat leaves and burrow holes in stems, fruit or roots. Moreover, they can transmit bacterial, viral or fungal infection to a crop. Effective insect control for farms and agricultural premises is essential.

Regular monitoring of agricultural lands can detect early invasion and using chemical spray like insecticides at larvae stage can indirectly reduce amount of insecticides needed. It is essential to control its outbreak. To do so, farmers need to scout their maize crop daily to be able to detect it early. Walking to the large fields and manually identify pests may take many hours.

Over the centuries, technologies have been applying to gain better yields. Intelligent robots and drones will revolutionise the efficiency of farms and change the agricultural landscape. These technologies can significantly reduce the amount of time needed for farmers and release them from manually monitoring their crops.

One of the main tasks for these technologies is to identify the presence of pests in the crops. Compared with the traditional machine learning methods, the diseases and pest classification method based on deep learning can directly input the original image. Instead of the tedious steps such as image pre-processing,

feature extraction and feature classification, the end-to-end structure is adopted to simplify the recognition process and solve the problem that the feature extractor designed manually is difficult to obtain the feature expression closest to the natural attribute of the object. The application of deep learning object detection can not only save time and effort, but also achieve real-time judgment to greatly reduce the huge loss caused by pests, which has important research value and significance.

However, pest classification on field crops, such as rice, soybean and other crops, is more challenging than generic object classification due to the inherent complexity of the background image. Throughout different stages of development of the insect (nymph or larva and adult), the morphological characteristics of insect vary greatly. To this end, we propose an efficient method to detect pests from the images captured on the fields. The proposed method can achieve a high accuracy in pest detection but with less amount of training data and time.

The rest of this paper is organised as follows: Section 2 reviews the related works of pest classifications using conventional machine learning methods and deep learning approaches. Then, Section 3 describes the proposed method and is followed by Experiments in Section 4. Section 5 finally concludes this paper summarising its main contributions with proposal of future works.

## 2. Related Works

### 2.1. Object Classification

Object classification is one of the most popular and emergent research topic in computer vision. Its task is to classify objects from different object categories. It can be useful for object detection and tracking in images and videos. In general, two approaches for object classification are Conventional Machine Learning (ML) and Deep Learning (DL) as visualising at Figure 1.

The conventional approach is to use well-established Computer Vision techniques to extract features such as Scale Invariant Feature Transform (SIFT) [1], Speeded Up Robust Features (SURF) [2], BRIEF [3], etc. These features are fed into a conventional ML method, e.g. Support Vector Machine (SVM) [4], Decision Tree (DT), Random Forest (RF), Logistic Regression (LR) etc. The decision of these methods will be a specific class. Choosing suitable features is a crucial task for these conventional approaches. It is, however, time consuming to manually extract and do multiple experiments to select appropriate features.

In contrast, Deep learning, also known as artificial neural network (ANN), consists of input layers, hidden layers and output layers. DL uses input layers to take an image dataset which each image is labelled as a specific

object class. Thereby, a DL model is trained on that data and it automatically discovers underlying patterns and generates feature maps for each specific class. This approach eliminates the need of handcrafted features.

Several DL architectures have been developed for object classification, e.g. LeNet [5], AlexNet [6], VGG16 [7], VGG19, Google Inception [8] and Resnet [9]. They achieved significant accuracy on the MNIST, ImageNet, PASCAL VOC2007.

However, these architectures are computational expensive. EfficientNet [10], a real time object classification, was published to address this issue. In general, authors of EfficientNet observed that in order to achieve better accuracy, the CNN is scaled up by increasing the number of layers, or making each layer wider, or inputting images at a higher resolution, or combining these factors. Figure 2 visualises these cases. However, the exploration of all these possibilities can be a tedious task and time consuming.

Neural Architecture Search (NAS) was utilised to develop a baseline architecture called EfficientNet-B0 described in Figure 3. Using the scaling search, EfficientNet-B0 is scaled up to EfficientNet-B1, and through EfficientNet-B7. The scaling function from one architecture to another is saved and applied to subsequent scalings because additional search becomes excessively expensive. EfficientNet performance has been compared with other popular object classifications on ImageNet, described on Figure 4. They achieve both higher accuracy and better efficiency over existing DL. Another important aspect is that they significantly reduce the number of parameters of the DL network.

### 2.2. Object Classification in Agriculture

Insects cause major damages to growing crops. For example, many insects can eat leaves or burrows in stems, fruit, or roots, while others can transmit bacterial, fungal infection into crops. When a small area is affected, it can easily spread out in a larger area. At this stage, it is too late and difficult for treatment and cost more. Therefore, early detection of insects is critical for application of different techniques in proper timing to control its outbreak. Proper timing is one of the best ways to control the pests. To do so, farmers need to scout their maize crop daily to be able to detect it early. Therefore, crop inspection is a vital part of farming. Walking to the large fields and manually identify pests may take many hours.

Several works have been developed to quickly identify insects from images with a certain accuracy due to various difficulties:

- Images captured from crops can contain a complex background.

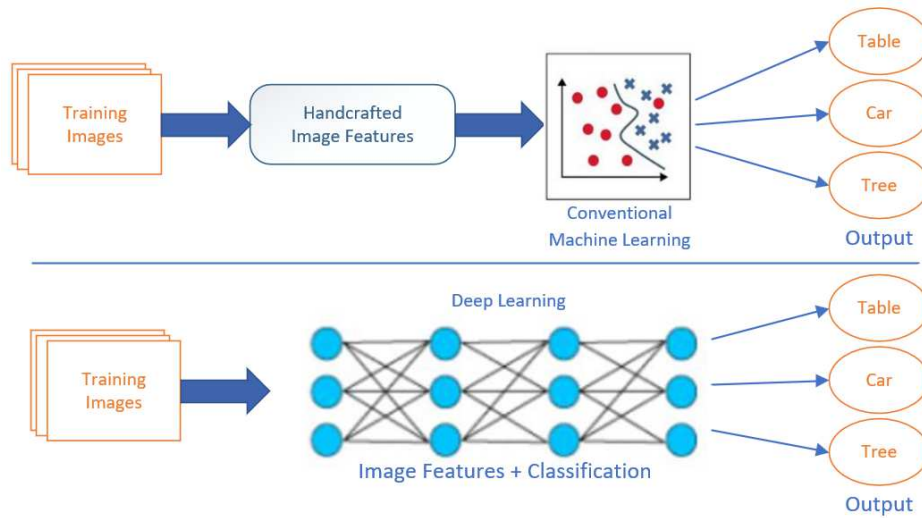


Figure 1. Two approaches for object classification: Conventional ML and DL

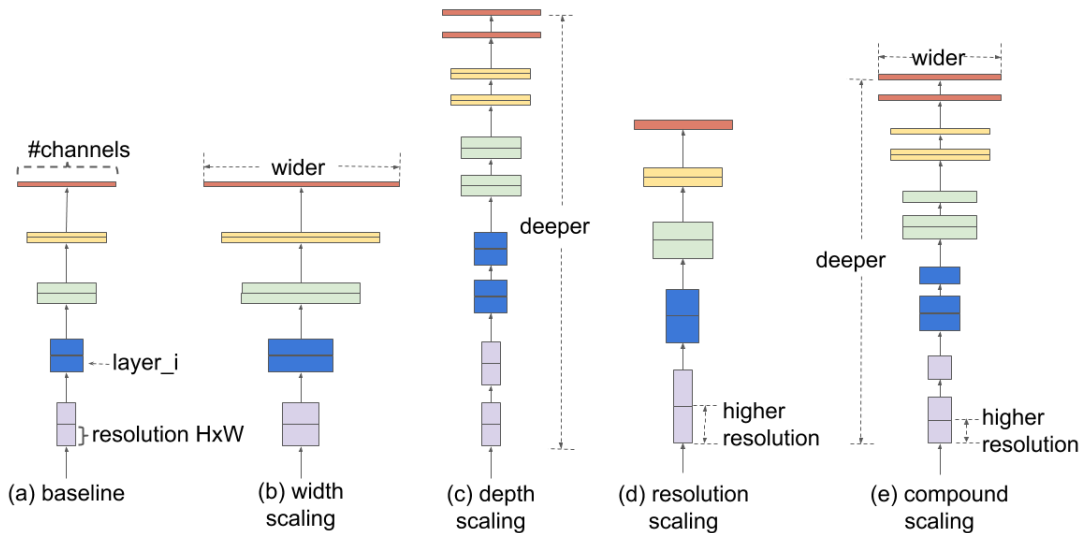


Figure 2. Model Scaling. (a) is an example of a baseline network; (b)–(d) are conventional scaling that only increases one dimension of network, width, depth or resolution. (e) uniformly scales all three dimensions with a fixed ratio. [10]

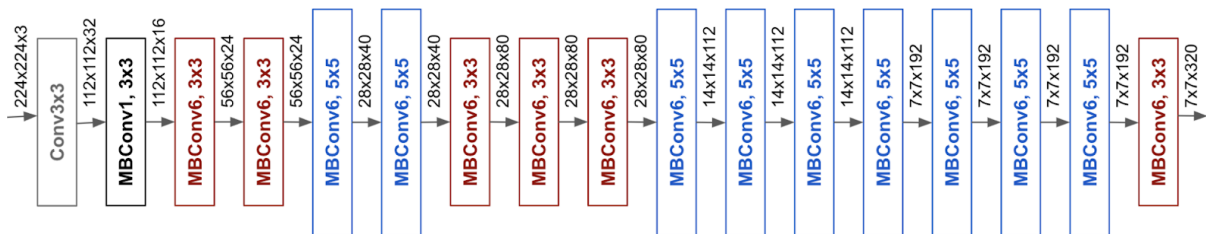
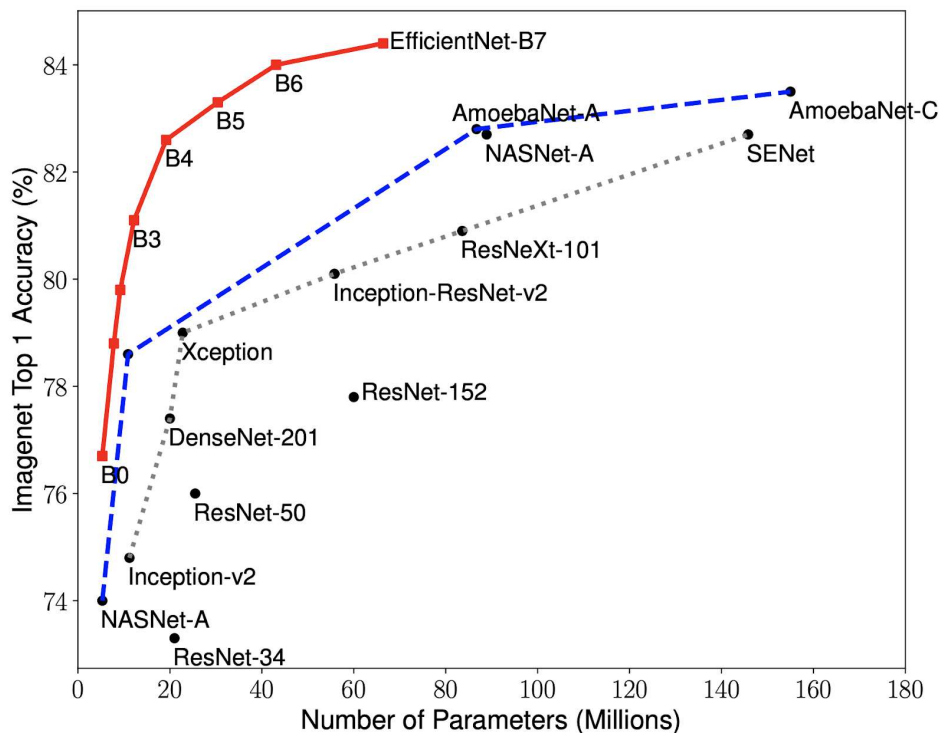


Figure 3. The EfficientNet-B0 architecture as a baseline network [10].

- Insects can have similar appearance as other objects in crops.
- There are different forms of the same insect species.
- Insects can be covered by leaves.

Xie et al. [11] developed multiple task sparse representation of insect objects. It combined a sparse-coding



**Figure 4.** Model Size vs. Accuracy Comparison. EfficientNet-B0 is the baseline network, while Efficient-B1 to B7 are obtained by scaling up the baseline network. [10]

technique and a multiple kernel learning technique as described in Figure 5. However, as mentioned above, this approach requires object segmentation and image denoising which are time consuming.

Ding and Taylor [12] proposed a DL based method to identify and counting codling moths in images captured inside field traps. The DL architecture was adopted from Lenet5 [5] to classify each local patch image if it contains moth or not. This architecture worked fine in a controlled setting and requiring pre-processing images.

The next section discusses our method to classify multiple types of insects from outdoor images, i.e. images contain more complex background and insects are various in appearances, sizes, angles and positions.

### 3. Proposed Method

Conventional machine learning methods have been designed to solely tackle problems. However, in deep learning, transfer learning [13] is an approach where a model was first trained on one dataset. It is similar to how human can utilise their knowledge across domain. In transfer learning, several layers of the trained model are used in a new model. The new model will be trained on another dataset. The benefit of transfer learning is to reduce the training time for a DL. The weights from pre-train layers are used as the initial weights for the

training process and will be adapted during the new training.

Figure 6 shows how the proposed DL architecture built from pre-train EfficientNet. The transfer learning with fine tuning is described by following steps:

1. Load a pre-trained model a neural network model. We use pre-train models EfficientNet on trained on the ImageNet as a baseline.
2. Develop a new DL architecture which replicates the designs and parameters of the pre-train model except the output layer.
3. Add an output layer with randomly initialised parameters for the new DL model with an output size equal to the number of target data set categories. ReLu is an activate function for the output layer and drop out rate is set to 75%.
4. Apply the Cross entropy as a loss function
5. Train the new DL model on the new data set. Only the new output layer is trained from the scratch. All remaining layers of the new DL model are freed. During the training process, the early stop method is exploited when the validation loss does not improve.

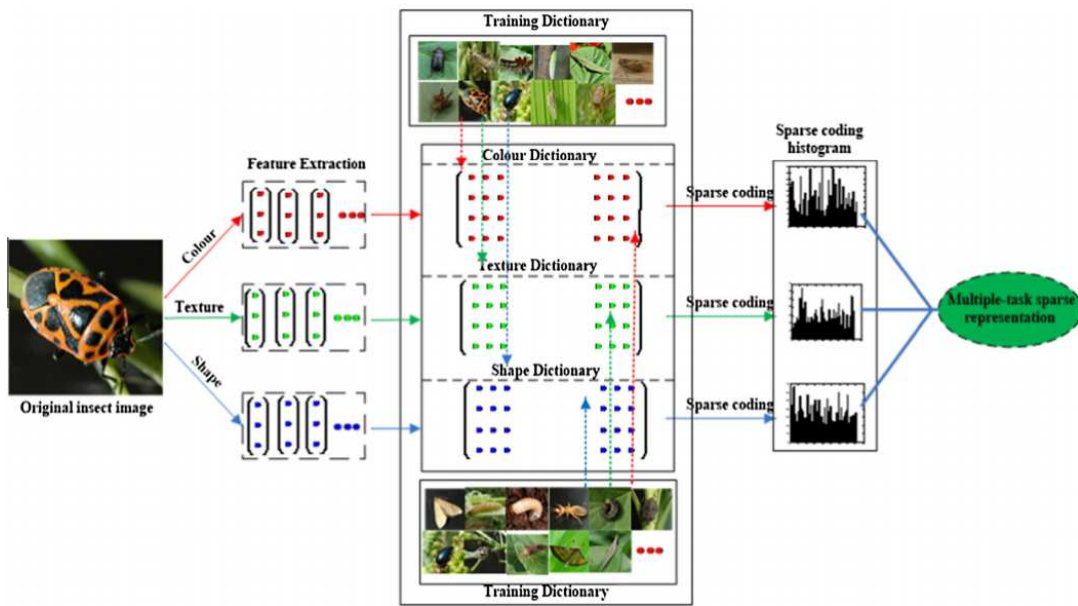


Figure 5. Multiple task sparse representation for Insect appearance [11].

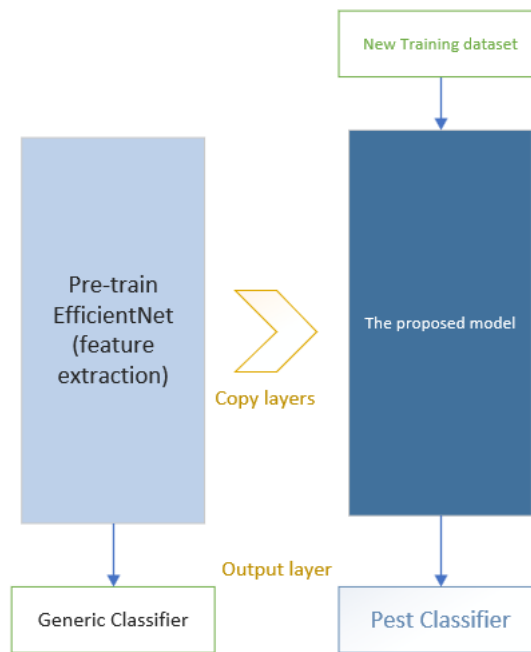


Figure 6. The proposed DL architecture, fine tuned from EfficientNet.

## 4. Experiments

### 4.1. Dataset

The insect dataset in this work contains 4,449 images and is downloaded from Kaggle. It contains 5 categories: Laybird, Mosquito, Grasshopper, Butterfly and Dragonfly. Table 1 shows the total number of images for each category and Figure 7 shows some examples of insect images. It is a challenge dataset because images are various sizes, contain complex

backgrounds and insects with different appearances, positions and poses.

Table 1. Number of images per insect category

Ladybird	Mosquito	Grasshopper	Butterfly	Dragonfly
864	690	960	899	1036



Figure 7. Some insect images extracted from the dataset.

## 4.2. Data preparation & augmentation

Each category in the dataset is splitted into 90% training data and 10% testing data. 20% of testing data is used for the validation step during the training process. The model will be trained on the training dataset and fine tuned by the validation dataset.

Next, data preparation involves using techniques such as the normalisation to rescale input data prior to training a DL model. Various range of data input may increase the error gradient values and make the learning process slow or unstable. Therefore, it is a good practice to have a normalised input data. In our work, before normalising all training data, each image is resized to 224 x 224. This dimension was used in popular DL network like AlexNet, VGG16 etc.

Enormous data are needed to improve a DL model. Collecting data is, however, difficult and time consuming. Therefore, data augmentation is a technique to artificially generate new training data from existing training data. Another benefit of data augmentation is to eliminate overfitting problem of the model. In this work, we randomly crop images with the size 224 x 224 from the original images. The generated images can be randomly rotated by 10 degree. They can be flipped around the horizontal axis.

## 4.3. Experiment Setup

Our proposed DL network is implemented by using pytorch. Its hyper parameters for training DL model are set with the batch size 8 and total epochs 10. The learning rate is set 0.000744 and scheduled to decrease 20% for each epoch during training to allow more fine-grained weight updates. Section 4.5.2.1 shows how to select the starting learning rate and total epochs selected.

It is, then, trained with CUDA 10.2 and cuDNN as the back-ends on GPU Tesla P100. During the training, we implement early stop technique to avoid overfitting. The training will stop if the accuracy is not improved after 3 times.

## 4.4. Evaluation metrics

The most common method to evaluate the performance of object classification is to calculate Precision (P) at Eq.1, Recall (R) at Eq.2, Accuracy (A) at Eq.3, F-measure (F) at Eq.4. To calculate these factors, a confusion matrix is needed as Figure 8 to calculate True positive (TP), False Negative (FN), False Positive (FP) and True Negative (TN) for each category.

$$\text{Precision}(P) = \frac{TP}{TP + FP} \quad (1)$$

$$\text{Recall}(R) = \frac{TP}{TP + FN} \quad (2)$$

$$\text{Accuracy}(A) = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)$$

$$F - \text{Measure}(F) = 2 \times \frac{P \times R}{P + R} \quad (4)$$

## 4.5. Result & Discussion

**4.5.1. Result.** We implemented 8 various models of EfficientNet to evaluate their performance. Table 2 summaries the performance for each model. The numbers in bold are the highest values.

Table 2. Accuracy of 8 models. The values in bold are the highest scores in respective columns.

Model	Precision	Recall	Acc	F-Measure
EfficientNet-B0	92.34	92.47	92.44	92.29
EfficientNet-B1	94.60	94.59	94.40	94.54
EfficientNet-B2	95.14	95.15	95.24	95.13
EfficientNet-B3	<b>95.49</b>	95.66	<b>95.52</b>	<b>95.54</b>
EfficientNet-B4	95.31	<b>95.77</b>	<b>95.52</b>	95.46
EfficientNet-B5	94.91	95.00	94.96	94.91
EfficientNet-B6	95.35	95.45	95.23	95.35
EfficientNet-B7	94.19	94.33	94.12	94.21

**4.5.2. Discussion.** Our experiments showed that EfficientNet-B3 achieves the best accuracy 95.52%. Below sessions discuss how to select an optimum learning rate and numbers of epochs.

		Predicted value	
		Positive	Negative
Actual value	Positive	True Positive	False Negative
	Negative	False Positive	True Negative

Figure 8. A general confusion matrix

**4.5.2.1. Effects of Learning Rate** In our implementation, we used Adam optimiser as a stochastic gradient descent method. The learning rate is needed for this method. When the learning rate is low, the training is reliable. However, the optimisation takes more time. Whereas if the learning rate is high, the training may not converge or diverge. We adopted the method explained in [14] to estimate the optimum learning rate which can be used as the starting point. Figure 9 shows the recommended learning rate is 0.000744.

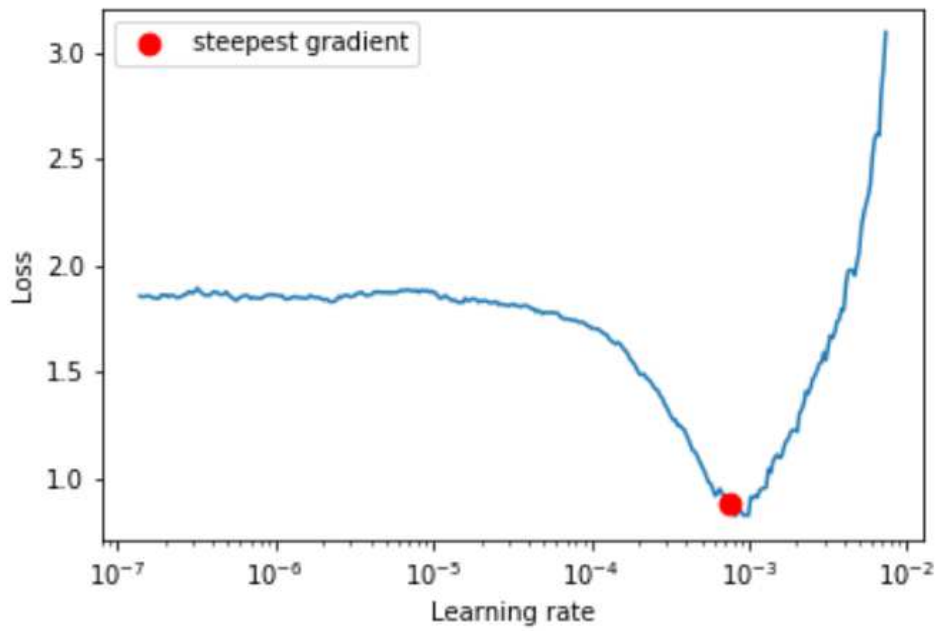
**4.5.2.2. Effects of number epochs** In deep learning, one epoch is one cycle when the whole training dataset is used. One epoch is divided into several small batches because the vast amount of the data. Training a deep learning model needs more than one epoch to pass the whole dataset several times to the same deep learning model. There is no absolute solution to identify which are the right numbers of epoch. Therefore, we implemented the early stop during the training process if there is no improvement of the loss. We observed that the maximum epochs in our proposed methods are 10 as displayed in Figure 10. The figure shows that after there is no significant improvement after 10 epochs.

## 5. Conclusions

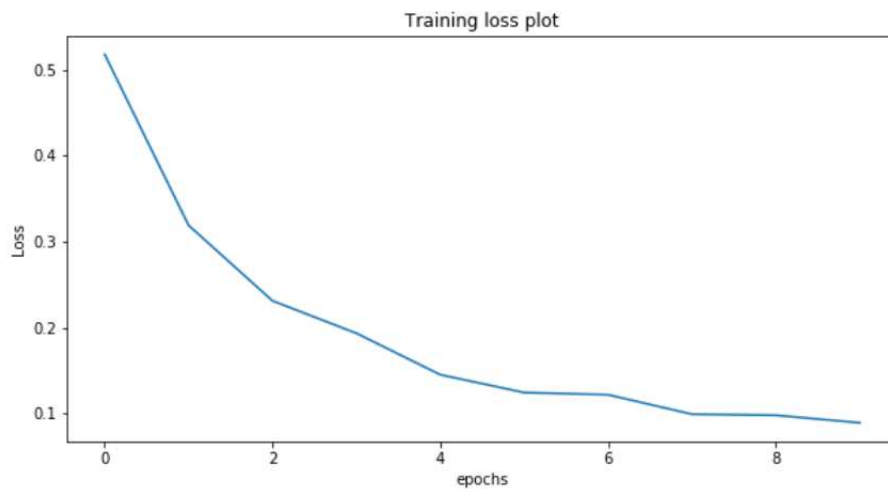
Pest classification is an important aspect in a smart agriculture. In this paper, we have proposed an efficient method basing on EfficientNet to classify pests from images with complex background and various pests in different forms. Specifically, we have evaluated 8 architectures of EfficientNet. All models were trained within 10 epochs but they achieved a significantly high accuracy, greater than 95%, for all criteria, Precision, Recall, Accuracy and F1-measure. In our experiment, EfficientNet-b3 is shown to achieve the best performance. In future, this proposed method can be experimented on embedded systems for autonomous machines like drones or robots used in agriculture.

## References

- [1] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vision*, vol. 60, no. 2, p. 91–110, Nov. 2004. [Online]. Available: <https://doi.org/10.1023/B:VISL.0000029664.99615.94>
- [2] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (surf)," *Comput. Vis. Image Underst.*, vol. 110, no. 3, p. 346–359, Jun. 2008. [Online]. Available: <https://doi.org/10.1016/j.cviu.2007.09.014>
- [3] M. Calonder, V. Lepetit, M. Ozuysal, T. Trzcinski, C. Strecha, and P. Fua, "Brief: Computing a local binary descriptor very fast," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 7, pp. 1281–1298, 2012.
- [4] M. A. Hearst, S. T. Dumais, E. Osuna, J. Platt, and B. Scholkopf, "Support vector machines," *IEEE Intelligent Systems and their Applications*, vol. 13, no. 4, pp. 18–28, 1998.
- [5] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.
- [6] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems*, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds., vol. 25. Curran Associates, Inc., 2012, pp. 1097–1105.
- [7] S. Liu and W. Deng, "Very deep convolutional neural network based image classification using small training sample size," in *2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR)*, 2015, pp. 730–734.
- [8] C. Szegedy, Wei Liu, Yangqing Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015.
- [9] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.
- [10] M. Tan and Q. V. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," 2020.
- [11] C. Xie, J. Zhang, R. Li, J. Li, P. Hong, J. Xia, and P. Chen, "Automatic classification for field crop insects via multiple-task sparse representation and multiple-kernel learning," *Computers and Electronics in Agriculture*, 2015.
- [12] W. Ding and G. Taylor, "Automatic moth detection from trap images for pest management," *Computers and Electronics in Agriculture*, 2016.
- [13] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning - Adaptive computation and machine learning*. MIT Press, 2016.
- [14] L. N. Smith, "Cyclical learning rates for training neural networks," in *2017 IEEE Winter Conference on*



**Figure 9.** The loss decreases in the beginning, then the training process starts diverging. It shows the recommended the optimum learning rate for the current training.



**Figure 10.** The change of loss values vs. numbers of epochs during the training.

*Applications of Computer Vision (WACV)*, 2017, pp. 464–472.